2.1 Å Crystal Structure of the Human Papillomavirus Type 18 E2 Activation Domain

S. F. Harris, E.A. Abbate, M.R. Botchan
Department of Molecular and Cell Biology
University of California, Berkeley, California 94720-3204, USA

INTRODUCTION

Papillomaviruses establish infection in dividing stem cells of various epithelial tissue. Certain high risk variants of human papillomavirus (HPV), including HPV18, are associated with cervical cancer and other intraepithelial neoplasias. The E2 protein is a key regulator of the virus' life cycle, controlling aspects of viral transcription, replication, and DNA segregation. The 45 kD E2 protein has a ~200 amino acid NH_2-terminal activation domain (AD) and a ~100 residue COOH-terminal DNA binding domain (DBD) separated by a flexible, protease-sensitive linker. E2 dimerizes and binds DNA at specific sites as seen in the solved DBD crystal structure of the bovine variant[1]. The activation domain, however, like activation domains in general, has remained enigmatic from a structural standpoint although mutagenesis studies have shown it to contain residues that are critical in each of E2's functions. Using multi-wavelength anomalous diffraction data, we have solved the crystal structure of a core region of the activation domain of HPV18 E2[2]. This, one of the first high-resolution views of such a domain, broadens the concept of activators to include proteins with potentially malleable, but certainly ordered structures. The E2 activation domain's novel fold creates a bent cashew-shaped form. A long α helix and a broken helical turn pack against a β sheet framework while connecting loops are generally exposed at the edges of the protein. Key determinants of E2's replication and transcription functions are widespread and often overlapping on the protein surface, though some small patches appear more dedicated to one or another of these activities. We further characterized one such region on the face of the amino terminal helix using mutagenesis and functional assays. Our studies suggest possible means of interaction for this protein and will help guide further experiments to elucidate fundamental aspects of transcription and replication mechanisms.

RESULTS

We performed limited proteolysis to identify structured cores of the protein. Digestion of the HPV18 E2 variant produced two nested activation domain fragments, identified by mass spectrometry and amino terminal sequencing as corresponding to residues 1-215 and 66-215. Crystals were eventually grown of both constructs, though those of the shorter fragment produced better diffraction. Native data to 2.1 Å was collected on these crystals at the Advanced Light Source (ALS) beamline 5.0.2. Selenomethionine-substituted crystals permitted the application of multiwavelength anomalous diffraction (MAD) techniques[3] to determine phases. 4-wavelength MAD data sets to 2.8 Å were collected at the ALS and at Stanford Synchrotron Radiation Labs (SSRL) beamline 1-5. Electron density maps were calculated from these phased data, and density modification improved the maps to allow the construction of a protein model including all but 7 of the possible residues in the crystallized construct.

A ribbon diagram and molecular surface of the solved structure are shown in figure 1. Given the high degree of sequence and functional conservation among the various viral E2 variants, we mapped the results of prior mutagenesis studies[4-8] onto the HPV18 E2 AD structure. Indeed, there is significant overlap of determinants of transcription and replication functions. This suggests that some architectural motifs may be shared in replication and transcription complexes.

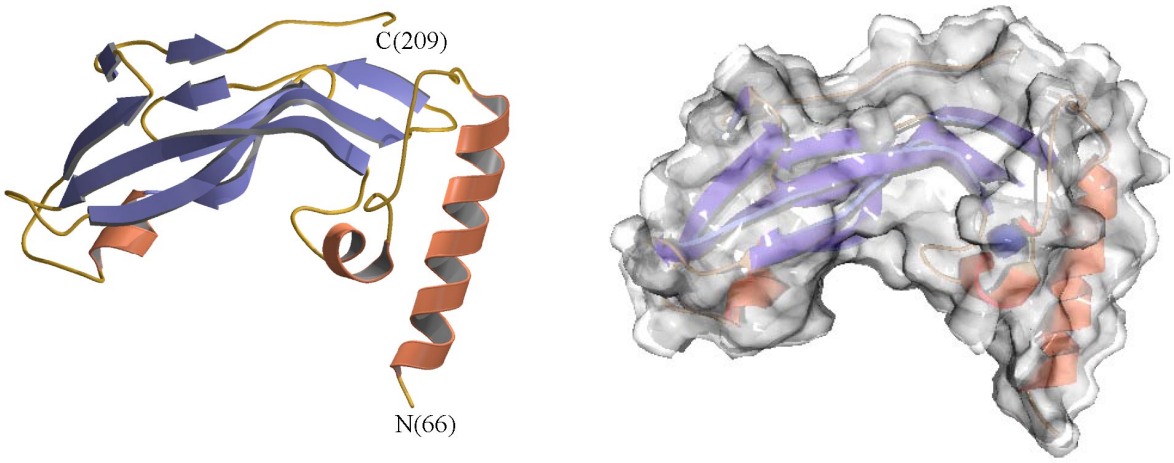

Figure 1. Ribbon diagram and molecular surface of the solved E2AD core.

At a finer level of detail, however, these two functions can be separated. Mutagenesis studies on several different E2 types showed that changing Ile73 to Ala abrogated transcription activation while leaving replication function essentially intact, and substituting Glu39 with Ala had the inverse phenotype, as this mutant E2 protein was no longer capable of stimulating viral replication[4-9]. We confirmed both phenotypes for the HPV18 variant. Ile73, although a hydrophobic residue, lies on the outer face of the long NH_2-terminal helix, suggestive of a possible role in protein:protein interaction. We targeted this helical face for further study, independently changing each of three exposed Gln residues to Ala (Figure 2). These, along with a double mutant of Gln76Ala/Gln84Ala and Glu39Ala, were tested for ability to stimulate transcription and replication in cell-based assays. Mutation of Gln76 shows a similar phenotype as its neighbor, Ile73: a strong effect on transcription activation. Further along the helix Ala substitution had lessening impact (Figure 2). All of these mutants maintained near wild type levels of replication stimulation, in contrast to mutation at site 39. The solved structure facilitated this more precise definition of a region dedicated to transcription function.

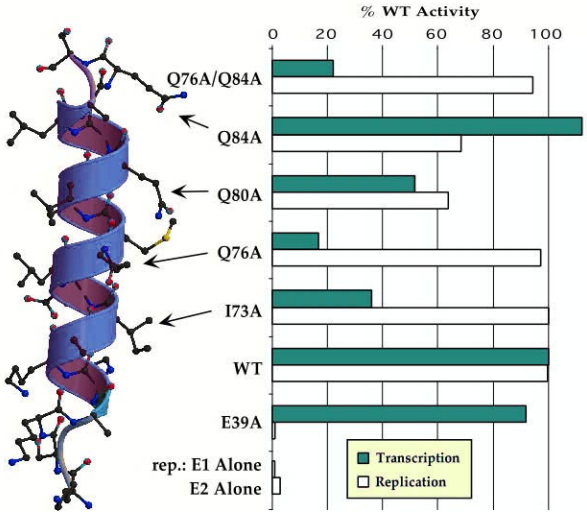

Figure 2. Transcription and replication effects of mutating residues on the outer face of E2's NH_2-terminal α helix.

An interesting correlation hints at possible mechanisms through which E2 may achieve transcription activation. In the crystal packing arrangements, another E2 molecule is positioned so that the long helices from two molecules are juxtaposed (Figure 3). Interactions between these regions are mediated primarily by Ile73 and Gln76, namely those residues that also appear most critical for transcription activation. A general model of transcription suggests that activators function by interacting with other members of transcription complexes, whether accessory factors or the polymerase itself, recruiting these proteins to appropriate sites on the DNA template. The exposed helical face of E2 may mediate such interactions either directly, or

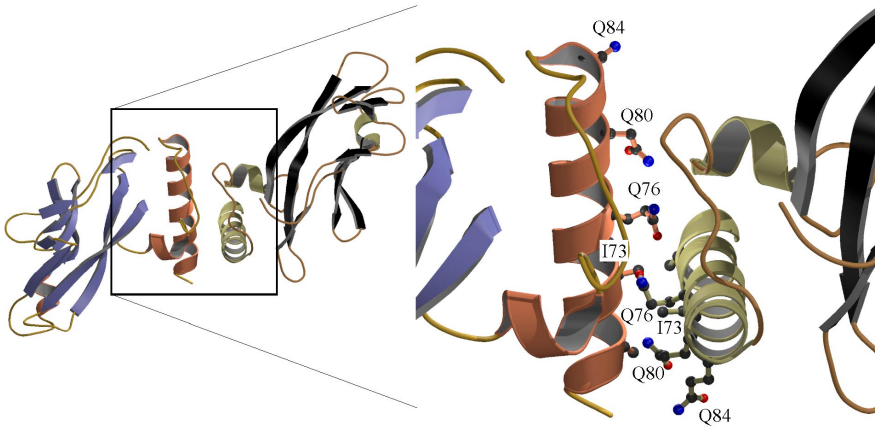

Figure 3. Crystal packing arrangements show intermolecular interaction between helical faces of E2 molecules, the same regions whose mutation affects transcription function.

perhaps through dimerization of E2 (as seen in the crystal symmetry) that then forms a composite surface necessary for recognition by transcription factors.

Prior studies on small peptides from other activators have also shown α helical structures that have amphipathic character similar to this helix in E2, although those peptides were only ordered upon interaction with other proteins. The extant E2AD structure has demonstrated the possibility that activators in isolation can be ordered, while corroborating the idea of amphipathic helices as a structural motif in transcription activation. Such biological insights will continue to be gleaned from the features of the E2AD structure. We expect it to guide further mutagenesis for a more detailed exploration of the replication and segregation functions of the protein. We are working to collect sufficient data on crystals of the longer construct which includes additional predicted NH_2-terminal helices that contain residues important for E2's interaction with the viral helicase E1.

ACKNOWLEDGMENTS

We are grateful to B. Spiller for much help in crystallography aspects of the work. We would like to think D. King for mass spectrometry work and the staffs at the SSRL and the ALS, in particular T. Earnest and L-W. Hung for their aid and time.

REFERENCES

1. R. S. Hegde, S. R. Grossman, L. A. Laimins and P. B. Sigler (1992). Nature *359*, 505-12.
2. S. F. Harris and M. R. Botchan (1999). Science *284*, 1673-7.
3. W. A. Hendrickson (1991). Science *254*, 51-8.
4. A. Abroi, R. Kurg and M. Ustav (1996). J Virol *70*, 6169-79.
5. D. E. Breiding, M. J. Grossel and E. J. Androphy (1996). Virology *221*, 34-43.
6. J. L. Brokaw, M. Blanco and A. A. McBride (1996). J Virol *70*, 23-9.
7. M. K. Ferguson and M. R. Botchan (1996). J Virol *70*, 4193-9.
8. H. Sakai, T. Yasugi, J. D. Benson, J. J. Dowhanick and P. M. Howley (1996). J Virol *70*, 1602-11.
9. F. Stubenrauch, A. M. Colbert and L. A. Laimins (1998). J Virol *72*, 8115-23.

This work was supported by NIH grants CA30490 and CA42414. The structure coordinates are available in the Protein Data Bank under code 1QQH.

Principal investigator: Michael Botchan, Department of Molecular and Cell Biology, University of California, Berkeley. Email: mbotchan@uclink4.berkeley.edu. Telephone: 510-642-7057.